



De Risi, R., De Luca, F., Gilder, C., Pokhrel, R., & Vardanega, P. J. (2020). The SAFER Geodatabase for the Kathmandu Valley: Bayesian Kriging for data-scarce regions. *Earthquake Spectra*.
<https://doi.org/10.1177/8755293020970977>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1177/8755293020970977](https://doi.org/10.1177/8755293020970977)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Earthquake Engineering Research Institute at <https://doi.org/10.1177/8755293020970977> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

The SAFER geodatabase for the Kathmandu valley: Bayesian kriging for data-scarce regions

Earthquake Spectra

1–19

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/8755293020970977

journals.sagepub.com/home/eqs

Raffaele De Risi, M.EERI^{ID}, Flavia De Luca, M.EERI^{ID},
Charlotte EL Gilder, M.EERI^{ID}, Rama Mohan Pokhrel^{ID},
and Paul J Vardanega^{ID}

Abstract

Geostatistical methods are valuable to better understand the spatial distribution of geotechnical parameters at regional scale and to optimize the locations of future ground investigations. This article investigates the use of the kriging interpolation method to extend the knowledge of a specific geotechnical property from a few sites to a broader geographical area with a focus on the Kathmandu valley (Nepal). A Bayesian form of kriging is proposed in this article. The estimation of the shear wave velocity in the uppermost 30 m of soil (V_{s30}) in the Kathmandu valley is examined. Slope-based V_{s30} estimates from the United States Geological Survey are used as prior information, and 15 V_{s30} measurements are used as more precise data. Considering the limited number of high-quality V_{s30} measurements available in the valley, it is shown that the Bayesian scheme can lead to a more robust estimation of V_{s30} than that obtained with the ordinary kriging approach. A methodology for conditioning prior low-precision data to the measurements is also presented.

Keywords

Kathmandu valley, Bayesian kriging, V_{s30} , SAFER geodatabase, soil classification

Date received: 21 March 2019; accepted: 25 September 2020

Department of Civil Engineering, University of Bristol, Bristol, UK

Corresponding author:

Flavia De Luca, Department of Civil Engineering, University of Bristol, Queens Building, University Walk, Bristol BS8 1TR, UK.

Email: flavia.deluca@bristol.ac.uk

Introduction

The Kathmandu valley in Nepal experienced significant seismic motions during the M_w 7.8 Gorkha earthquake that occurred on 25 April 2015 (Goda et al., 2015). Such ground motions were exceptionally high due to the amplification caused by the soil conditions (e.g. Rajaure et al., 2017; Tallett-Williams et al., 2016), which led to significant widespread structural and geotechnical damage (e.g. Goda et al., 2015; McGowan et al., 2017). Seismic motions at a given location are mainly influenced by the characteristics of the seismic source and the wave path (McGuire, 2004). Seismic source characteristics include the epicenter/hypocenter location, the faulting style, fault geometry, and the earthquake magnitude. Typical wave path parameters are the distance between the site of interest and the seismic source, and the site response (e.g. De Risi et al., 2019). The study of the site response effects can be performed with varying levels of sophistication (International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE), 1999). The most common approach uses soil classification (Stewart et al., 2014) which is based either on the shear wave velocity in the uppermost 30 m of soil (V_{S30}) (Foti et al., 2018) or proxy information, for example, local topography or geo-lithology (Allen and Wald, 2009; Wald and Allen, 2007). In either case, a reliable database of geotechnical properties is required. V_{S30} is commonly used in design codes such as Eurocode 8 (CEN, 2004) or ASCE 7-10 (ASCE, 2010) and in ground motion prediction equations (GMPEs; Douglas, 2003) to estimate the site amplification.

The compilation of a V_{S30} database is expensive and time-consuming. It thus presents a challenge due to (a) the limited amount of financial resources typically devoted to geotechnical testing, which invariably permits few measurements over a large geographical area, and (b) the potential lack of a repository where the data acquired over time are stored in a systematic, coherent, and accessible way. Geotechnical properties are usually individual point measurements located in readily accessible areas that are not necessarily representative of the mapped geologic unit (Thompson et al., 2014). The interpolation from individual observations to a larger geographical area for mapping purposes is of paramount importance for geotechnical engineering decision-making, especially in data-scarce regions (Rahman et al., 2018). To make the most of available data and inform acquisition of new data in a cost-effective manner, new algorithms for data acquisition/analysis are needed.

This article presents a new Bayesian algorithm, based on the kriging interpolation approach (Chilès and Delfiner, 2012), that makes use of several layers of information at increasing quality for the creation of a spatial V_{S30} map. This can provide an informative distribution of V_{S30} over a wider geographical area than that covered by the input data. Traditionally, kriging has been used to combine different geotechnical data sources in order to obtain more accurate and reliable V_{S30} maps (Thompson et al., 2014), site amplification factor maps (Thompson et al., 2010), liquefaction potential maps (Pokhrel et al., 2013), or other geotechnical parameters (Marache et al., 2009). Kriging interpolation, in combination with the Bayesian approach, has been reported in the literature (Pilz and Spöck, 2008) and is known as Bayesian kriging (Omre, 1987; Omre and Halvorsen, 1989). The Bayesian approach is particularly useful for studying the site response variation (Chakraborty and Goto, 2018). In this article, the Bayesian form of kriging is mainly devoted to the improvement of the knowledge about the spatial variability of V_{S30} for the Kathmandu valley (Cui et al., 1995). Specifically, with respect to traditional kriging, the proposed approach allows for the quantification and propagation of the uncertainties on the parameters that define the spatial modeling of V_{S30} . As for V_{S30} , many parameters resulting from geological or geotechnical interpretations and measurements can be

considered non-stationary, as geological attributes rarely exist in a homogeneous domain. Conventional forms of kriging are able to deal with some, but not all, of the non-stationary effects of a particular attribute (Machuca-Mory and Deutsch, 2013). Yet, Bayesian kriging (a) allows the restrictions of unbiased prediction of conventional approaches to be overcome, (b) provides an accurate prediction of moderately non-stationary data, and, most importantly, (c) allows modeling standard errors of predictions with incremental accuracy. This is especially useful for small datasets such as the V_{S30} measurements available for the Kathmandu valley.

The kriging interpolation uses the variogram to quantify the spatial correlation (Chilès and Delfiner, 2012; Webster and Oliver, 2007); such a function is governed by few parameters that require sufficient observations to be stable and reliable. In this study, the variogram is the focus of the Bayesian algorithm. Specifically, the uncertain parameters of the variogram are determined on the basis of the freely available V_{S30} estimates provided by the United States Geological Survey (USGS; Wald and Allen, 2007). Such V_{S30} values are obtained as a function of geomorphological data (i.e. slope), which is used as a proxy for the local site characteristics with a resolution of 30 arcsec. Once the variability of the variogram's parameters is understood for each parameter, probabilistic models are fitted and are used as priors. Then, the direct measurements acquired by in-situ tests are used as the input to the likelihood function. These direct measurements of shear wave velocity are presented in Gilder et al. (2020). According to the basic hypotheses of kriging, the likelihood function is a joint log-normal distribution. The central values of the likelihood function are defined by the USGS slope-based estimates at the locations of the in-situ tests, and the covariance matrix of the likelihood function is calculated based on the variogram. By applying Bayes' theorem, it is possible to obtain the posterior distribution of the variogram's parameters and, therefore, a new and more precise estimation of the expected shear wave velocity values.

Several novel features are presented in this article. The first main novelty consists of the manner in which the prior distribution of the parameters of the variogram using the USGS V_{S30} data is created. Second, the Bayesian updating of the distributions of the parameters of the variogram and their propagation with the robust approach is presented here for the first time with respect to the case of the V_{S30} estimation within the kriging framework. Finally, the conditioning of the USGS map on the observations and the robust approach taking advantage of the availability of the prior and posterior distributions of the variogram parameters are new procedures.

Methodology

In this section, the ordinary and Bayesian kriging procedures are outlined; these procedures allow for the estimation of values of a specific geotechnical property (in this case, the V_{S30}) over larger areas at several unsampled points using few point measurements.

Ordinary kriging

Let $Z(\mathbf{u})$ be the generic random variable representative of a geotechnical property for one or more unsampled locations (\mathbf{u}), and let $z(\mathbf{u}^*)$ be the realization of the random variable Z representing the measured geotechnical property at the sampling locations (\mathbf{u}^*). For the case of geotechnical parameters, the random variable Z can be replaced with the natural logarithm of a geotechnical property, for example, $Z = \ln(V_{S30})$. The estimations of Z at the unsampled locations \mathbf{u} , that is, the interpolated values based on N observations, can be calculated as:

$$\hat{Z}(\mathbf{u}) = \sum_{i=1}^N \lambda_i \cdot z(\mathbf{u}_i^*) \quad (1)$$

where λ_i are the weights that, to make the estimation unbiased, need to sum to 1 and are obtained by minimizing the estimation variance:

$$\text{Var}[\hat{Z}(\mathbf{u})] = 2 \cdot \sum_{i=1}^N \lambda_i \cdot \gamma(\mathbf{u}_i^*, \mathbf{u}) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \cdot \lambda_j \cdot \gamma(\mathbf{u}_i^*, \mathbf{u}_j^*) \quad (2)$$

where $\gamma(\mathbf{u}_i^*, \mathbf{u})$ is the semi-variance between the i th data point (\mathbf{u}_i^*) and the target data point (\mathbf{u}), and $\gamma(\mathbf{u}_i^*, \mathbf{u}_j^*)$ is the semi-variance between the data points \mathbf{u}_i^* and \mathbf{u}_j^* . The variance defined in Equation 2 is also known as kriging variance, and it is equal to zero at the observation points and takes positive values elsewhere. The variance, being a function of γ , is also a function of the lag distance (h) between two points; the functional form between the variance and h is known as a variogram. Such a function allows for consideration of the spatial correlation, that takes into account the way a specific random variable varies in space.

The variogram can be bounded or unbounded and can take different analytical forms (Chilès and Delfiner, 2012; Webster and Oliver, 2007). The main parameters of the variogram are the *nugget*, the *range*, and the *sill*. The nugget is the value of the function for a lag distance equal to zero; the range is the lag distance for which the function becomes constant (i.e. the autocorrelation becomes zero); and the sill is the maximum value of the variogram and represents the maximum variance of the process. In this study, a spherical variogram function is adopted since it is among the most adopted for isotropic variables (Webster and Oliver, 2007):

$$\gamma(h) = \begin{cases} c \cdot \left[\frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \cdot \left(\frac{h}{a} \right)^3 \right] & \text{for } h \leq a \\ c & \text{for } h > a \end{cases} \quad (3)$$

where a is the range, c is the sill, and the nugget is equal to zero.

Bayesian kriging

Figure 1 illustrates the main steps of the Bayesian kriging approach used in this article for a hypothetical problem domain. Let $\boldsymbol{\theta}$ represent the parameters of the variogram (i.e. $\boldsymbol{\theta} = \{c, a\}$) that need to be estimated on the basis of newly available data \mathbf{D} . $\boldsymbol{\theta}$ may be considered as a set of random variables represented by (marginal/joint) probability distribution functions. According to the Bayesian paradigm, the distribution of $\boldsymbol{\theta}$ can be updated with new observations (Box and Tiao, 1992):

$$f(\boldsymbol{\theta}|\mathbf{D}) = C^{-1} \cdot L(\mathbf{D}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}) \quad (4)$$

where $f(\boldsymbol{\theta})$ is the prior distribution of the variogram parameters and identifies the available information on $\boldsymbol{\theta}$ prior to the estimation and prior to the acquisition of more reliable data \mathbf{D} (e.g. the V_{S30} geophysical measurements), $L(\mathbf{D}|\boldsymbol{\theta})$ is the likelihood function and characterizes the information from the newly available observations \mathbf{D} , $f(\boldsymbol{\theta}|\mathbf{D})$ is the posterior distribution describing the newly updated estimate of $\boldsymbol{\theta}$, and C is a normalization factor (De Risi et al., 2017). This approach allows the determination of the distributions of the

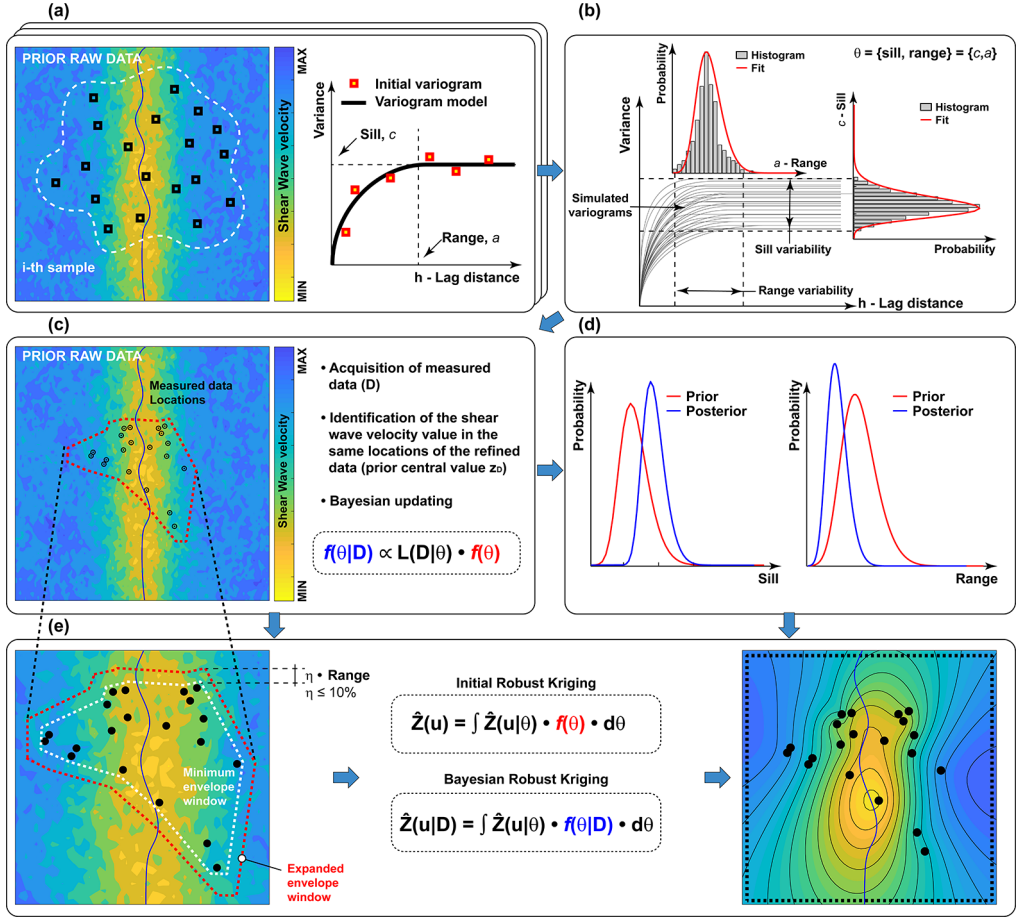


Figure 1. (a) Sampling of data and variogram fitting. (b) Definition of the prior distribution for sill and range. (c) Acquisition of more precise data and Bayesian updating setup. (d) Calculation of the posterior distribution of sill and range. (e) Initial and robust Bayesian kriging.

parameters of the variogram. These distributions will facilitate the construction of a Bayesian predictor to be used in the kriging approach according to Equations 1 and 2.

Informative priors $f(\theta)$ should be used in order to reflect the current state of knowledge about the parameters of interest (Pilz and Spöck, 2008). To elicit informative distributions, it is possible to use experience or prior data, that, in the early stages of the analysis, do not need to be highly informative. To obtain the prior distribution of the parameters of the variogram θ , a random sample of the prior data at random locations was performed; then, the natural logarithm of the samples is fitted with a variogram, and therefore the sill and range $\{c_i, a_i\}$ are obtained (Figure 1a). Repeating the sampling on the prior data many times allows a large number of variograms to be produced and hence a large number of potential values of sill and range. This sampling procedure is consistent with the so-called empirical Bayesian kriging (Krivoruchko and Gribov, 2014). The initial distributions of sill and range can be fitted with analytical distributions (Figure 1b) describing the prior knowledge of the parameters of the variogram $f(\theta)$.

If the generic random variable $Z(\mathbf{u})$ is spatially distributed as a Gaussian random field, the likelihood function in Equation 4 is a multivariate normal with mean $(z_D = z(\mathbf{u}_D^*))$ equal to the natural logarithm of the prior information at the locations (\mathbf{u}_D^*) of the new N data \mathbf{D} (Figure 1c), and the covariance matrix $(\Sigma_\gamma(\boldsymbol{\theta}))$ is the one obtained using the variogram function that depends on $\boldsymbol{\theta}$. Therefore, the likelihood function can be written as:

$$L(\mathbf{D}|\boldsymbol{\theta}) = (2\pi)^{-N/2} \cdot |\Sigma_\gamma|^{-1/2} \cdot e^{-\frac{1}{2}[\ln(\mathbf{D}) - z_D]' \cdot \Sigma_\gamma^{-1} \cdot [\ln(\mathbf{D}) - z_D]} \quad (5)$$

where the dependency of Σ_γ on $\boldsymbol{\theta}$ is omitted in the notation for simplicity. The covariance matrix is calculated for the natural logarithm of the random variable of interest.

From Equation 4, it is then possible to derive the posterior distribution of the variogram parameters $f(\boldsymbol{\theta}|\mathbf{D})$ (Figure 1d). Either prior or posterior distribution of the variogram parameters can be used to propagate the uncertainties about $\boldsymbol{\theta}$ in the final kriging estimation (Figure 1e). Depending on whether the prior or the posterior distribution of $\boldsymbol{\theta}$ is used, the initial robust kriging (IRK) or the Bayesian robust kriging (BRK) is calculated by Equations 6 and 7, respectively:

$$\hat{Z}(\mathbf{u}) = \int \hat{Z}(\mathbf{u}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta} \quad (6)$$

$$\hat{Z}(\mathbf{u}|\mathbf{D}) = \int \hat{Z}(\mathbf{u}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}|\mathbf{D}) \cdot d\boldsymbol{\theta} \quad (7)$$

where $\hat{Z}(\mathbf{u}|\boldsymbol{\theta})$ is calculated according to Equations 1 and 2 using as $z(\mathbf{u}^*)$ only the new data \mathbf{D} (i.e. direct measurements). The robustness of the estimation is due to the fact that the uncertainties on the parameters of the variogram are propagated completely on the final estimation taking advantage of the total probability theorem (Jalayer et al., 2015); i.e. $f(\boldsymbol{\theta})$ and $f(\boldsymbol{\theta}|\mathbf{D})$ represent the degree of confidence on $\boldsymbol{\theta}$ (i.e. the weight of $\boldsymbol{\theta}$), and the integrals of Equations 6 and 7 are a weighted average of the kriging estimation based on the direct measurements \mathbf{D} . In other words, the robust approach can be seen as a weighted average of kriging results where the weights are the probability density function (PDF) values of the parameters of the variogram.

Analogously to Equations 6 and 7, it is possible to compute maps of the robust kriging variance substituting $\hat{Z}(\mathbf{u}|\boldsymbol{\theta})$ with $\text{Var}[\hat{Z}(\mathbf{u}|\boldsymbol{\theta})]$ evaluated according to Equation 2:

$$\text{Var}[\hat{Z}(\mathbf{u})] = \int \text{Var}[\hat{Z}(\mathbf{u}|\boldsymbol{\theta})] \cdot f(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta} \quad (8)$$

$$\text{Var}[\hat{Z}(\mathbf{u}|\mathbf{D})] = \int \text{Var}[\hat{Z}(\mathbf{u}|\boldsymbol{\theta})] \cdot f(\boldsymbol{\theta}|\mathbf{D}) \cdot d\boldsymbol{\theta} \quad (9)$$

This variance quantifies, in a robust manner, the variability between the observations and the prediction. Furthermore, since the robust kriging is the expected value of \hat{Z} for all the potential values of the model's parameters, another estimation of the variance can be calculated as follows for the cases of IRK and BRK, respectively:

$$\sigma_R^2[\hat{Z}(\mathbf{u})] = \int \hat{Z}(\mathbf{u}|\boldsymbol{\theta})^2 \cdot f(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta} - \left[\int \hat{Z}(\mathbf{u}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta} \right]^2 \quad (10)$$

$$\sigma_R^2[\hat{Z}(\mathbf{u}|\mathbf{D})] = \int \hat{Z}(\mathbf{u}|\boldsymbol{\theta})^2 \cdot f(\boldsymbol{\theta}|\mathbf{D}) \cdot d\boldsymbol{\theta} - \left[\int \hat{Z}(\mathbf{u}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}|\mathbf{D}) \cdot d\boldsymbol{\theta} \right]^2 \quad (11)$$

This latter variance estimation provides the local variability associated with the robust assessment.

Conditioning the prior to measured data

The geographical coverage of the direct measurements \mathbf{D} , in general, are not as extensive as for the less precise prior data. Therefore, the estimation of $\hat{Z}(\mathbf{u})$ by Equation 6 or 7 is limited to the geographical coverage of \mathbf{D} (e.g. the minimum envelope represented in Figure 1e as dashed white line) in which only interpolation is performed, or to a slightly larger area (e.g. the expanded envelope window in Figure 1e represented in a dashed red line, obtained expanding the minimum envelope domain of a small percentage η of the range, for example, $\eta \leq 10\%$) for which both interpolation and extrapolation are needed. The limitation on the extrapolation area allows limiting the variability of the extrapolation. For example, to a value of η equal to 10% corresponds a variance equal to the 15% of the sill (Equation 3), allowing the control of this parameter in the extrapolation area by the user. However, the domain can be limited using any other physically informed rationale; for example, the extrapolation could be contained only within a specific geological feature (e.g. a basin within a mountain range with significantly different geological features).

Several approaches exist in the literature to anchor regional data to local observations (Miano et al., 2016; Worden et al., 2010). Building upon these methods, to extend the estimation to a larger geographical area, for example, the one covered by the prior data, it is possible to condition the prior data to the more precise new data \mathbf{D} according to the following expression (Eaton, 1983):

$$z(\mathbf{u}|\boldsymbol{\theta}) = z(\mathbf{u}) + \boldsymbol{\Sigma}_{\gamma, z_D \ln(\mathbf{D})} \cdot \boldsymbol{\Sigma}_{\gamma, \ln(\mathbf{D}) \ln(\mathbf{D})}^{-1} \cdot [\ln(\mathbf{D}) - z_D] \quad (12)$$

where $z(\mathbf{u})$ represents the logarithm of the prior data, $\boldsymbol{\Sigma}_{\gamma, z_D \ln(\mathbf{D})}$ is the cross-covariance matrix for the logarithm of the prior values in correspondence of the new data (z_D) and the logarithm of the new data ($\ln(\mathbf{D})$), and $\boldsymbol{\Sigma}_{\gamma, \ln(\mathbf{D}) \ln(\mathbf{D})}$ is the covariance of the logarithm of the new data. Both covariance matrices can be obtained using the variogram function for which the prior or posterior distributions of the parameters are employed.

This well-consolidated conditioning can be integrated and improved using the robust approach discussed previously. The estimates $z(\mathbf{u}|\boldsymbol{\theta})$ obtained from Equation 12 can be used in Equation 6 or 7 (using either prior or posterior distribution) to obtain robust conditioning of the initial estimates on the new measurements allowing to propagate the uncertainties of the variogram parameters into the conditioned map. Therefore, it is possible to obtain updated robust maps with a geographical extent identical to the prior data.

The Kathmandu valley case

In this study the shear wave velocities provided by the USGS V_{S30} model (Allen and Wald, 2009) were used as prior data. The V_{S30} map is shown in Figure 2a. The map describes well the general distribution of the metamorphic bedrock surrounding the valley and characterizes the inner distribution of valley sediments of lacustrine and fluvio-deltaic origins (Shrestha et al., 1998). Further information on the geotechnical parameters associated with these soils are provided in Gilder et al. (2020). On the same plot, the area of the valley sediments identified from the geological map is enclosed within the blue line, and an example of random sampling across

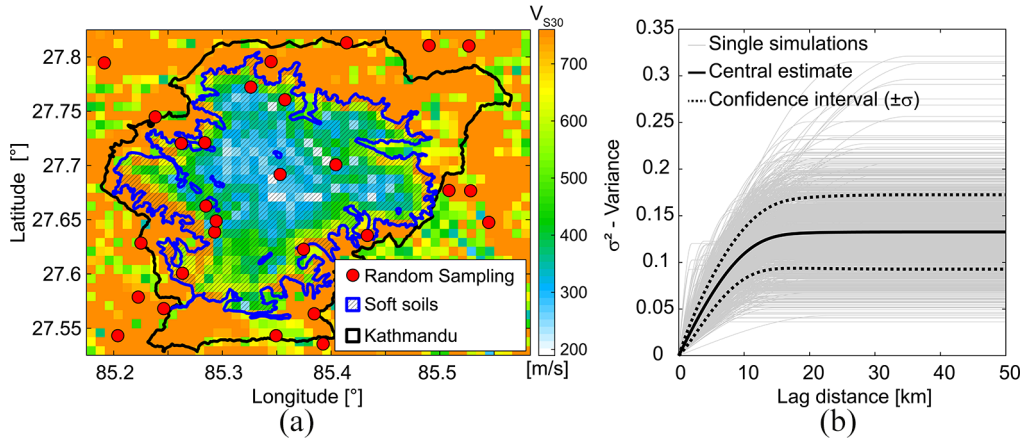


Figure 2. (a) USGS V_{S30} database for Kathmandu and example of random sampling. (b) Initial variogram functions; ($0.05^\circ \approx 5.6$ km).

the greater area is also shown. For this specific study, it was observed that 30 samplings are the minimum number to fit a variogram in a stable manner. Repeating the sampling 1000 times, 1000 variogram functions are generated (gray lines in Figure 2b) from which it has been possible to obtain empirical distributions of range and sill values.

To investigate the sensitivity of the sampling to the dimension of the sampling domain, two different areas were investigated: (a) the larger area corresponding to the geographical extent in Figure 2a, and (b) a smaller area corresponding to the area of sediments, enclosed in the blue domain in Figure 2a. Figure 3 shows the distributions of sill and range for both the larger area (Figure 3a and b) and the soft soil area (Figure 3c and d). For both the c and a parameters of the variograms, a generalized extreme value (GEV) distribution (Kotz and Nadarajah, 2000) was found to be the model that best fit the data compared to other simple models (e.g. Log-Normal, Normal, Weibull). In this work, sill and range are assumed to be uncorrelated. Only a small reduction in the central value of the range and the dispersion can be observed between the distributions obtained for the two sampling domains. The fitted distributions can be used as informative priors in Equation 4. Figure 3a and c, as well as Figure 3b and d, shows very similar results. Although very similar, these two priors are used here to check whether the posterior is sensitive to the extent of the sampling domain. The triplets of parameters presented in parenthesis in Figure 3 are the parameters of the GEV, that is, the shape factor p_1 , the scale factor p_2 , and the location parameter p_3 , respectively. Equation 13 shows the analytical expression of the GEV PDF for a generic variable x :

$$f(x) = \left(\frac{1}{p_2}\right) \exp \left[- \left(1 + p_1 \frac{x - p_3}{p_2} \right)^{-\frac{1}{p_1}} \right] \left(1 + p_1 \frac{x - p_3}{p_2} \right)^{-1 - \frac{1}{p_1}} \quad (13)$$

As presented in Gilder et al. (2020), 18 direct geophysical measurements of V_{S30} are available in the Kathmandu valley (Figure 4a), also see Gilder et al. (2019) for the database SAFER/GEO-591. Since some measurements are in the same location, they are averaged and considered only once. Therefore, 15 measured V_{S30} were available for the analysis presented in this article. Table 1 lists the V_{S30} values obtained from both the downhole seismic tests and the USGS model. The USGS values tend to systematically overestimate the

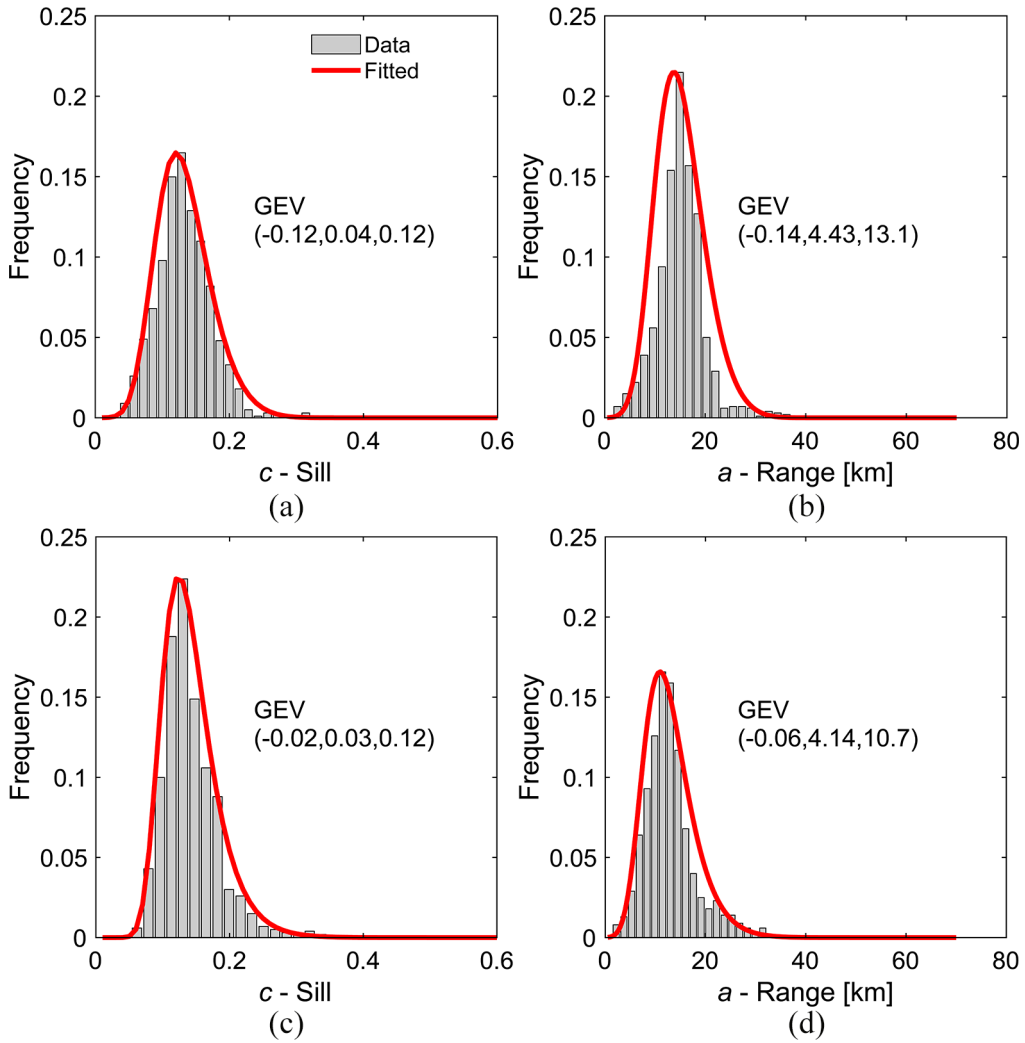


Figure 3. (a, c) Distributions of the sill. (b, d) Distributions of the range. Results for the (a, b) greater area and (c, d) sediment area.

available borehole data for the Kathmandu valley, especially for the boreholes close to steep locations (e.g. B11), as also observed in Stewart et al. (2014). Moreover, as observed by Allen and Wald (2009), the lower elevations provided the worst estimates. Some measured V_{S30} values are particularly low with respect to the USGS estimations (e.g. B6, B8, B10, B11, B12); these sites correspond to locations that are underlain by Quaternary (i.e. recent deposits or Talus Cone deposits), as well as the Plio-Pleistocene sediments.

Initially, the measured V_{S30} values are used in the ordinary kriging (Figure 4b and c) and to compute the standard error $\left(\sigma = \sqrt{\text{Var}[\hat{Z}(\mathbf{u})]}\right)$ associated with the geostatistical analysis (Figure 4d). These results are presented here for comparison with the results presented later in the article. As discussed earlier, the error is minimum in correspondence of the measurement location and increases moving away from these measurements. The error

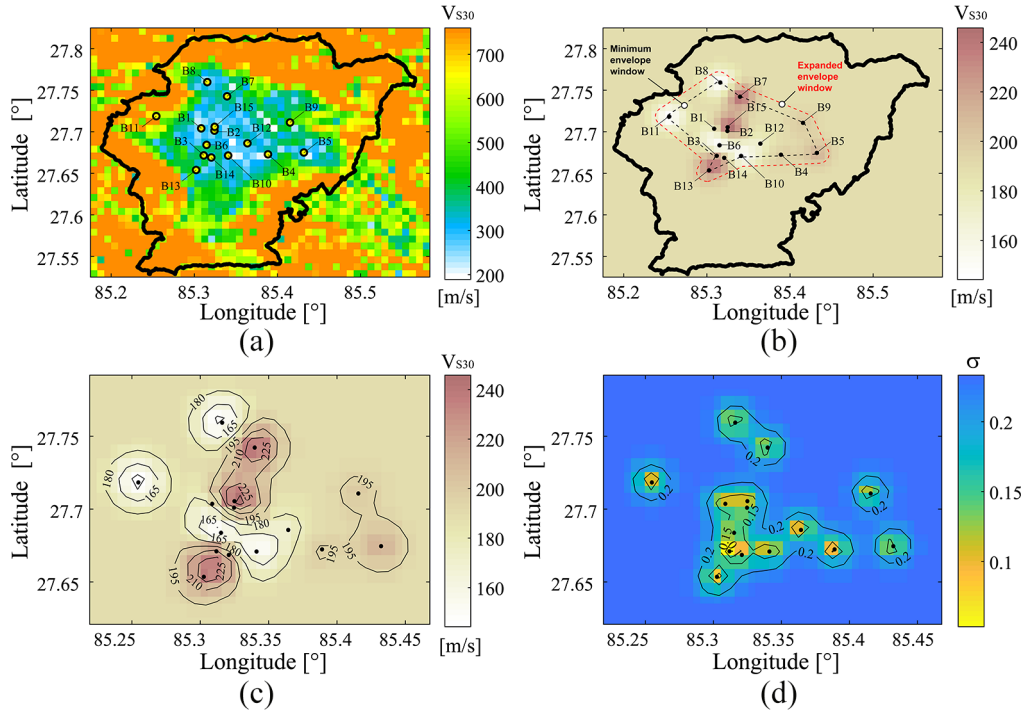


Figure 4. (a) Location of the measurements. (b) Ordinary kriging based on the measurements and variogram. (c) Ordinary kriging for the small window. (d) Kriging standard deviation; ($0.05^\circ \approx 5.6$ km).

Table 1. V_{s30} at the 15 locations from direct geophysical measurements (*Geophysical Measurements*), from slope-based estimation by USGS (*USGS*) and the logarithm of the ratio between the two sources of information ($\ln(\text{Measured}/\text{USGS})$)

ID	Database ID	Location		V_{s30} (m/s)		\ln (Measured/USGS)
		Longitude (°)	Latitude (°)	Measured	USGS	
B1	R_JICA_2002_BH1	85.3087	27.7036	180	279	-0.438
B2	R_JICA_2002_BH2	85.3247	27.7009	231	295	-0.245
B3	R_JICA_2002_BH3	85.3118	27.671	219	314	-0.360
B4	R_JICA_2002_BH4	85.3891	27.6724	198	250	-0.233
B5	R_JICA_2002_BH5	85.4322	27.6745	216	278	-0.252
B6	IND_Bakh_2006_ BH1, BH3	85.3152	27.6837	140	343	-0.893
B7	IND_Bans_2007_BH3, BH5 & BH8	85.3398	27.7422	254	341	-0.295
B8	R_JRAP_2016_BH1	85.316	27.7593	140	263	-0.631
B9	R_JRAP_2016_BH2	85.4156	27.7107	203	391	-0.656
B10	R_JRAP_2016_BH3	85.3412	27.6709	147	264	-0.586
B11	R_JRAP_2016_BH4	85.2547	27.7183	139	606	-1.472
B12	R_JRAP_2016_BH5	85.3643	27.6857	170	289	-0.531
B13	RES_POKH_2006_BH6	85.3026	27.6535	237	419	-0.570
B14	RES_POKH_2006_BH7	85.3209	27.6686	207	307	-0.394
B15	RES_Safe_2018_BH1	85.325	27.7054	257	290	-0.121

USGS: United States Geological Survey.

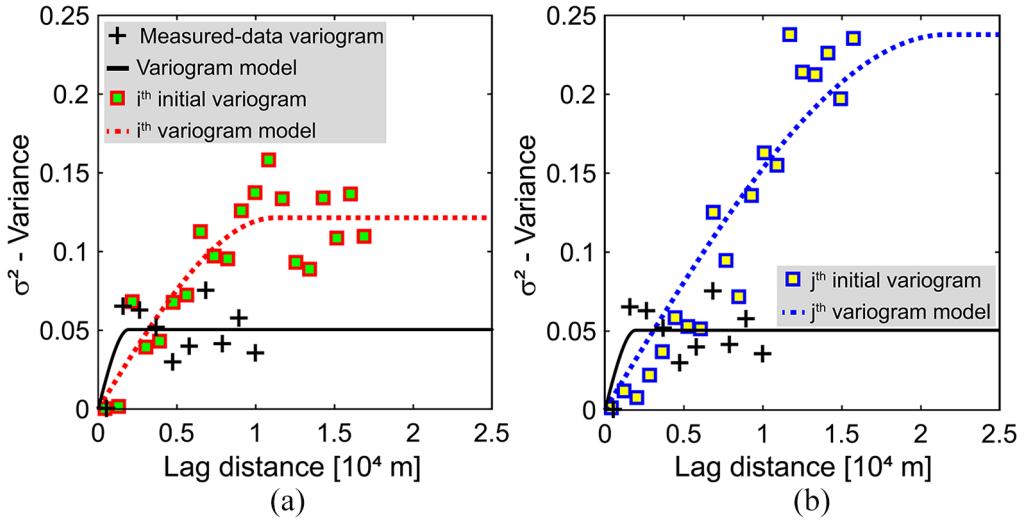


Figure 5. Comparison of the variograms calculated using the measured data and two random samplings from the USGS map.

map allows the identification of sites for future field investigations (e.g. the blue area in Figure 4d). However, it is not possible to account only for the reduction in the prediction error; a risk-based approach should be used. Further constraints must be considered (e.g. the exposure, the seismic vulnerability of the assets at stake). Hence, identifying locations for future geotechnical field investigations should be done to optimize the inevitable compromise between the error reduction (in a geostatistical sense) and the need of precise geotechnical data in areas where the risk may be highest (e.g. Gilder et al., 2018).

Figure 5 shows the difference between the variograms obtained using both the measured data (i.e. black cross markers and black lines) and the two generic random samplings from the USGS maps (i.e. the square markers and dashed lines in Figure 5a and 5b). As shown in Figure 2b, the two simulations in Figure 5a and b are selected to show the large variability that can be obtained in terms of initial variograms and its fit. The plots emphasize the significant difference that exists between a variogram generated according to the proposed sampling procedure (i.e. dashed curves in Figure 5a and b, respectively) and the one obtained according to a straightforward ordinary procedure (i.e. the black solid curve in Figure 5).

The direct geophysical measurements can be used in the likelihood function defined by Equation 5. Integrating Equation 5 and the prior distributions of the parameters of the variogram, it is possible to obtain the posterior distributions of the parameters of the variogram (Figure 6).

To demonstrate the influence of the prior distribution, results are shown for both the larger geographical (Figure 6a and b) and the sediment (Figure 6c and d) sampling domains. Figure 6a and c shows that the distribution of the sill shifts toward higher values, its variability remains constant, and the shape of the function tends to lose the positive skewness becoming more symmetric for the larger geographical domain meanwhile retaining the right skewness for the sediment area sampling domain. Figure 6b and d shows that the distribution of the range shifts to higher values for the domain containing

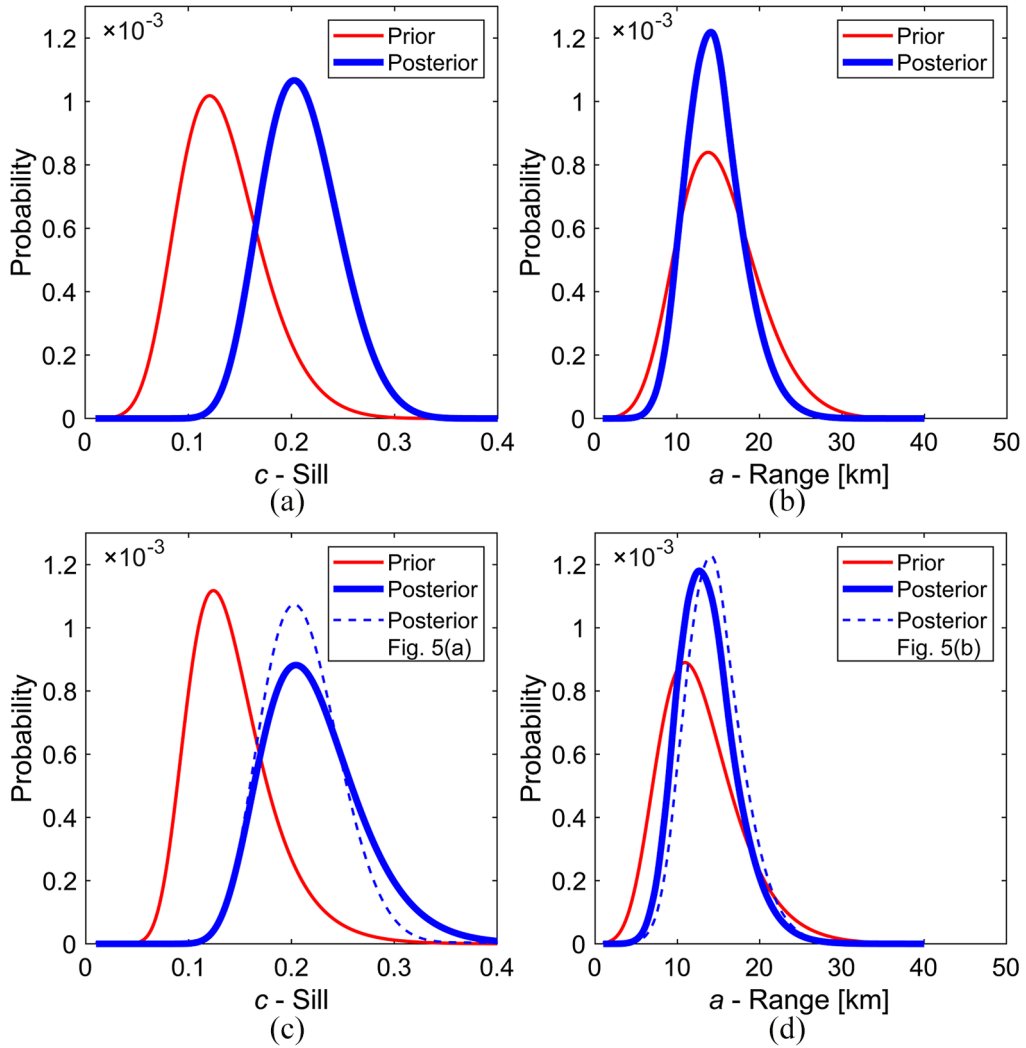


Figure 6. (a, c) Prior and posterior distributions of the sill. (b, d) Prior and posterior distributions of the range. Results for the (a, b) larger area (entire Kathmandu Valley) and (c, d) sediment area (soft soil area) as per Figure 2a.

sediments only. Nevertheless, for both sampling domains, variability is lower, and skewness is toward higher values of the range. The results presented in Figure 6 demonstrate that the maps obtained using the posterior parameters will have a larger variability for the same amount of lag distance. Therefore, posterior maps will present a more gradual variation of estimates and will reflect the fact that the V_{S30} is not purely distributed according to the topographical features. Moreover, it can be observed that the posteriors obtained using the two sampling domains are similar (continuous and dashed blue lines in Figure 6c and d). Both the prior and the posterior distributions of the parameters of the variogram can be used in Equations 6 and 7 to obtain robust kriging estimations. Figure 7a and c shows the IRK, and Figure 7b and d shows the BRK. In addition, minimum and expanded envelope domains are shown in gray and red dashed lines. Regarding the comparison of IRK and BRK, the results are very similar. However, for both IRK and BRK

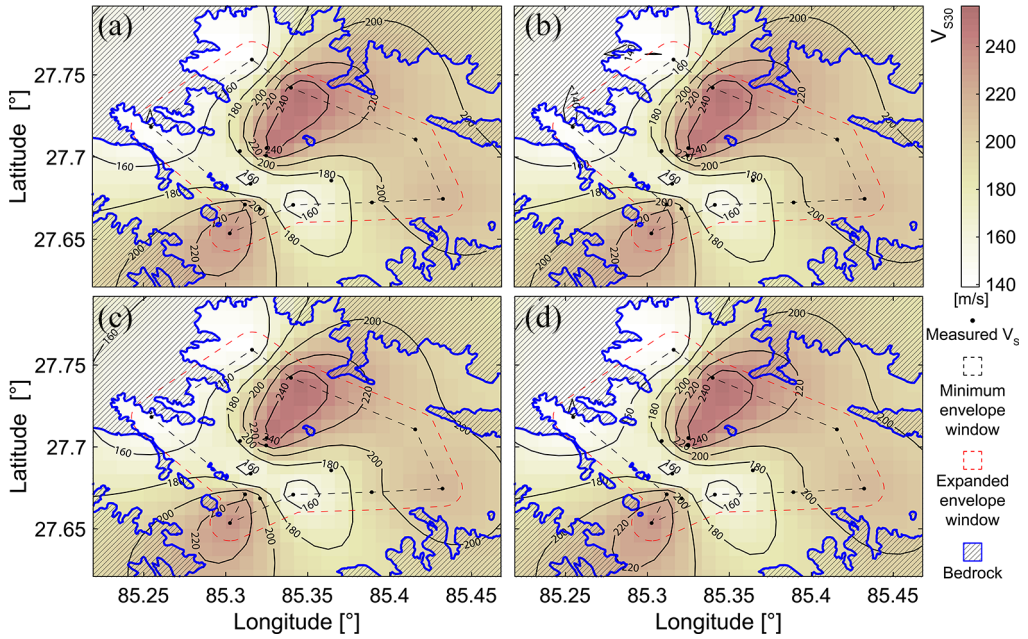


Figure 7. (a, c) Initial robust kriging (IRK) and (b, d) Bayesian robust kriging (BRK). Results for the (a, b) larger area and (c, d) soft soil area; ($0.05^\circ \approx 5.6$ km).

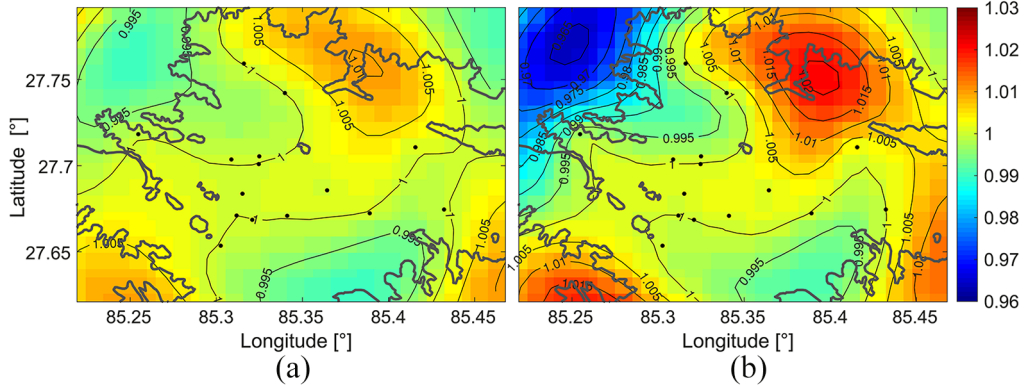


Figure 8. Ratio between BRK and IRK for the larger area and the soft soil area: (a) the results presented in Figure 7b and a. (b) the results presented in Figure 7d and c.

cases, the results are dissimilar from the results of the ordinary kriging presented in Figure 4c, especially with respect to the contour levels of V_{S30} .

Figure 7a and c, as well as Figure 7b and d, shows almost identical results. It is possible to conclude that the robust kriging, for the considered case study, is not sensitive to the shape of the priors of the variogram parameters. Therefore, in the following discussion, only the results obtained considering the larger sampling domain are presented. The larger sampling domain is preferred as it will allow a broader use for future updates of results from additional investigations done in the valley.

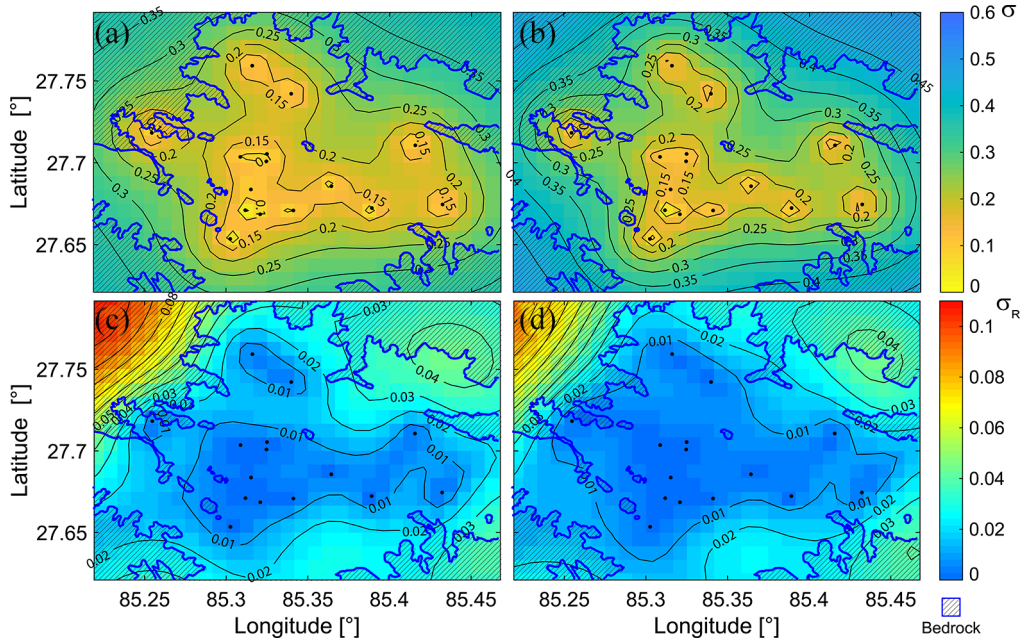


Figure 9. Kriging standard deviation for (a) initial robust kriging (IRK) as per Equation 8 and (b) Bayesian robust kriging (BRK) as per Equation 9. Local standard deviation for (c) IRK as per Equation 10 and (d) BRK as per Equation 11; ($0.05^\circ \approx 5.6$ km).

As expected, also high similarities between the results obtained with the posterior and the prior distributions of the parameters of the variograms can be observed. Figure 8 shows the ratio between the V_{S30} estimation obtained with the different hypotheses. It is possible to conclude that maximum variation of it is in the range of $\pm 3\%$ for this case study; therefore, the number of measured data is still too small to obtain significant variation with the Bayesian approach.

In addition to previous results, considering only the larger sampling domain, the maps of the standard deviation for the kriging estimation are also presented for both IRK and BRK. Specifically, Figure 9a and b shows the kriging variance obtained from Equations 8 and 9, respectively. Figure 9c and d shows the variance obtained from Equations 10 and 11, respectively. From Figure 9a and b, it is possible to conclude that the kriging variability for the BRK is slightly larger than the IRK one; therefore, although the estimation presented in Figure 7a and b is similar, the variability for the BRK is slightly larger, as expected looking at the posteriors of the sill. However, results presented in Figure 9c and d show that the weighted averaging estimation performed for the BRK leads to a lower local variability.

Conditional V_{S30} map

The 15 geophysical measurements can be used to condition prior data according to Equation 12. Figure 10a and b show the initial USGS V_{S30} map and the same map with conditioned V_{S30} values.

For the robust application, only the posterior distribution of the parameters of the variograms is used (i.e. the parameters distributed according to the blue curves in Figure 6a

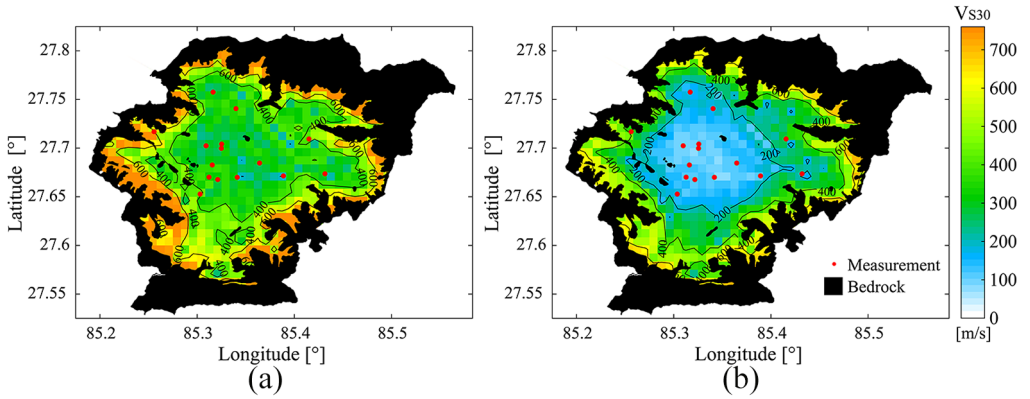


Figure 10. Shear wave velocity maps: (a) USGS estimates. (b) USGS estimates conditioned on the measurements; ($0.05^\circ \approx 5.6$ km).

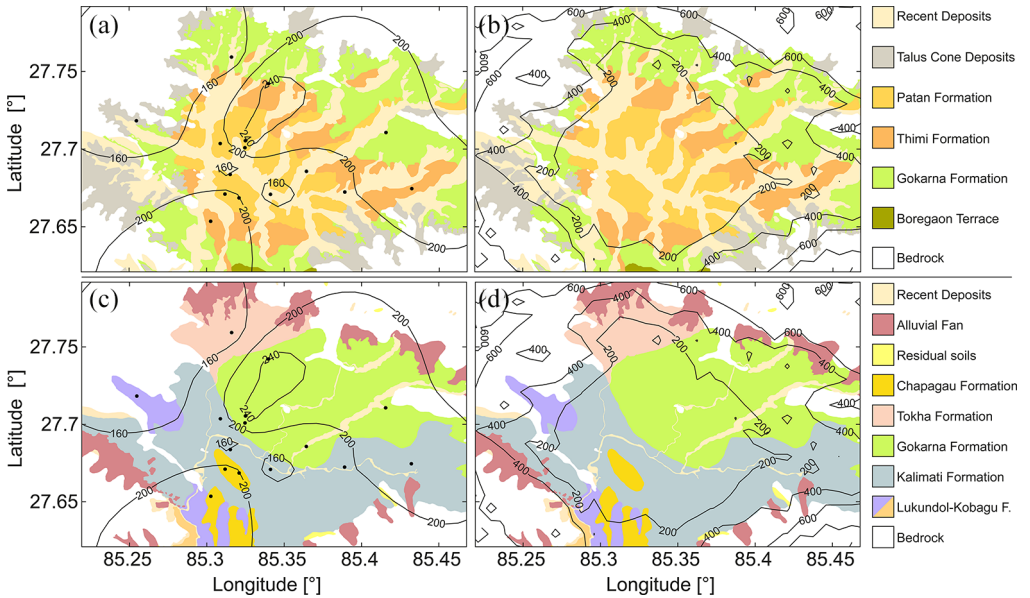


Figure 11. Overlay of the geology maps with the contours of the V_{S30} obtained considering (a, c) only the geophysical measurements and (b, d) the USGS estimates conditioned to the geophysical measurements. (a, b) Geology maps shown in Gilder et al. (2020) based on Yoshida and Igarashi (1984). (c, d) Geology maps shown in Gilder et al. (2020) based on Shrestha et al. (1998); ($0.05^\circ \approx 5.6$ km).

and b). The initial values are reduced by the measured values, and the distribution of the V_{S30} values in the central part of the valley is smoothed. The main effect on the result can be observed where there is a higher concentration of measurements. No significant variation of the V_{S30} values can be observed far from the locations of the measurements. In this case, the geologic bedrock is overlaid on the map (black area in Figure 10).

Comparisons

For comparison purposes, in Figure 11, the two contour maps of shear wave velocity obtained through BRK (Figure 7b) and USGS estimates conditioned to the measurements (Figure 10b) are compared with the geology maps given in the companion paper Gilder et al. (2020). The first map, originally developed by Yoshida and Igarashi (1984), provides information on the geology inferred from geomorphology (Figure 11a and b). The second map, developed by Shrestha et al. (1998), provides geological information as a result of engineering soil classification (Figure 11c and d).

Figure 11a shows how the interpolated results (i.e. in the minimum envelope domain) describe the variability of the shear wave velocities in the central part of the valley where the sediments are present very well. Figure 11b provides an update to the prior assumptions made in the USGS topographical model, and the geophysical data indicating the inner portion of the valley has even lower values than the global model, due to the specific deposits present in this area (both recent river deposits and tidal flat, coupled with an underlying silt lake deposit—Kalimati Formation). A critical aspect is realized when the engineering geological map is compared to the results, that is, sediment distributions based on origin and grain size. The overlay in Figure 11c shows that the kriging provides a distinction between the two main engineering soils present in the valley (Gokarna and Kalimati Formations). It may be expected that the Gokarna Formation will have higher values of V_{S30} , as it contains sand/gravel layers, interlayered within silt and clay. Alternatively, the Kalimati Formation (in the southern central valley) is a more homogeneous engineering material comprising a clay/silt which is expected to exhibit lower values of V_{S30} .

It is acknowledged that there remains a need for improvement of some aspects of the kriged distribution, as seen by the two boreholes in the upper-left part of Figure 11c; they are located in particularly soft/loose recent deposits, which is producing an imprecise estimation toward the bedrock, and potentially uninformative distribution across the remaining sediments in that portion of the valley. Where the bedrock might be expected to be reasonably shallow in these upper topographies, the V_{S30} data are not available. To summarize, Figure 11b and d showing the USGS values conditioned on the geophysical measurements can aid in providing lower values for a model spanning the entire valley. Alternatively, Figure 11a and c reflects the engineering soil classification, with local greater resolution information, very well.

Conclusions

This article details a new approach to extract controlled-confidence estimations from available geotechnical data in data-scarce regions. Ordinary kriging was used to draw maps representing the geographical variability of V_{S30} , which can aid decision in locating new geotechnical investigations. A Bayesian kriging approach has been proposed to combine several layers of data. Specifically, the slope-based V_{S30} estimates provided by the USGS have been used as prior, and direct geophysical measurements are used to inform the likelihood model. The posterior distributions obtained by means of the Bayesian approach can be used to obtain a robust kriging interpolation/extrapolation. Finally, using the posterior variograms, a procedure for conditioning the prior data to the geophysical measurements has been proposed, allowing extension of the estimations to a geographical coverage larger than that of the (more precise) direct estimations.

The procedures have been applied to the case study of the Kathmandu valley, Nepal, and the results have been compared with the geological information for the valley. Both ordinary and Bayesian kriging perform best in the central part of the valley, where the density of observations is larger; moreover, the proposed methodology provides more gradual change of the geotechnical property with respect to the ordinary kriging that instead works efficiently mainly around measurements and changes in a more abrupt manner. From the conditioned prior data, it has emerged that there is a substantial reduction in the initial values, and the distribution tends to change mainly where the concentration of observation is high providing suitable results for extrapolations. The progression of the increasing value of shear wave velocity from the center toward the mountains remains evident.

In the case study, the number of observations available does remain insufficient (i.e. 15 measurements) for such a large area to produce significantly different results between IRK and BRK. However, the novel approach proposed allows for full control of the confidence of the interpolation, and the robust approach allows for a characterization of the variance at the local level.

Several potential improvements should be explored in future work; for example, including the heterogeneity of the variables that were not discussed in this study. The sampling scheme adopted to create the priors can be modified considering a different technique. Moreover, variogram parameters could be considered jointly. Finally, advanced simulation routines may be used to solve the Bayesian problem.

Acknowledgments

The authors acknowledge the Engineering and Physical Science Research Council (EPSRC) project “Seismic Safety and Resilience of Schools in Nepal” SAFER (EP/P028926/1). The first author also acknowledges the EPSRC projects “Enhancing PREParedness for East African Countries through Seismic Resilience Engineering” PREPARE (EP/P028233/1) and SAFER PREPARED (EP/T015462/1). The third author acknowledges the support of EPSRC (EP/R51245X/1). The authors would like to thank the anonymous reviewers for a welcome exchange of ideas and for their contribution in improving the manuscript. The contour maps of V_{S30} obtained in this study (Figures 4b, 7a, 7b, and 10b) are provided in raster format as supplementary material. The database SAFER/GEO 591 is available for download from the University of Bristol Data Repository (Gilder et al., 2019).

Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding


The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Engineering and Physical Science Research Council (EPSRC) Grant codes: EP/P028926/1; EP/P028233/1, EP/R51245X/1.


ORCID iDs

Raffaele De Risi  <https://orcid.org/0000-0002-5496-9656>

Flavia De Luca  <https://orcid.org/0000-0003-2387-8580>

Charlotte E L Gilder  <https://orcid.org/0000-0002-7695-5490>

Rama Mohan Pokhrel  <https://orcid.org/0000-0002-9803-5556>

Paul J Vardanega  <https://orcid.org/0000-0001-7177-7851>

Supplemental material

Supplemental material for this article is available online.

References

- Allen TI and Wald DJ (2009) On the use of high-resolution topographic data as a proxy for seismic site conditions (VS 30). *Bulletin of the Seismological Society of America* 99(2A): 935–943.
- ASCE (2010) *Minimum Design Loads for Buildings and Other Structures*. Reston, VA: ASCE.
- Box GEP and Tiao GC (1992) *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons, pp. 588.
- CEN (2004) *Eurocode 8: Design of Structures for Earthquake Resistance—Part 1: General Rules, Seismic Actions and Rules for Buildings*. Brussels: CEN.
- Chakraborty A and Goto H (2018) A Bayesian model reflecting uncertainties on map resolutions with application to the study of site response variation. *Geophysical Journal International* 214(3): 2264–2276.
- Chilès J-P and Delfiner P (2012) *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. Hoboken, NJ: John Wiley & Sons, pp. 699.
- Cui H, Stein A and Myers DE (1995) Extension of spatial information, Bayesian kriging and updating of prior variogram parameters. *Environmetrics* 6(4): 373–384.
- De Risi R, Goda K, Mori N and Yasuda T (2017) Bayesian tsunami fragility modeling considering input data uncertainty. *Stochastic Environmental Research and Risk Assessment* 31(5): 1253–1269.
- De Risi R, Penna A and Simonelli LA (2019) Seismic risk at Urban scale: The role of site response analysis. *Soil Dynamics and Earthquake Engineering* 123: 320–336.
- Douglas J (2003) Earthquake ground motion estimation using strong-motion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews* 61(1–2): 43–104.
- Eaton ML (1983) *Multivariate Statistics: A Vector Space Approach*. New York: John Wiley & Sons, pp. 116–117.
- Foti S, Hollender F, Garofalo F, Albarello D, Asten M, Bard PY, Comina C, Cornou C, Cox B, Di Giulio G, Forbriger T, Hayashi K, Lunedei E, Martin A, Mercierat D, Ohrnberger M, Poggi V, Renailier F, Scilia D and Socco V (2018) Guidelines for the good practice of surface wave analysis: A product of the InterPACIFIC project. *Bulletin of Earthquake Engineering* 16(6): 2367–2420.
- Gilder C, De Risi R, De Luca F, Vardanega P, Holcombe EA, Ayoubi P, Asimaki D, Pokhrel RM and Sextos A (2018) Optimising resolution and improvement strategies for emerging geodatabases in developing countries. In: *Proceedings of the 16th European conference on earthquake engineering*, Thessaloniki, Greece, 18–21 June, Paper No. 10743. *Thessaloniki*: European Association for Earthquake Engineering.
- Gilder CEL, Pokhrel RM and Vardanega PJ (2019) *The SAFER Borehole Database (SAFER/GEO-591_v1.1)*. Bristol: University of Bristol, <https://doi.org/10.5523/bris.3gjcvx51lnpuv269xsa1yrb0rw>
- Gilder CEL, Pokhrel RM, Vardanega PJ, De Luca F, De Risi R, Werner M, Domniki A, Maksey PN and Sextos A (2020) The SAFER geodatabase for the Kathmandu valley: Geotechnical and geological variability. *Earthquake Spectra* 36(3): 1549–1569.
- Goda K, Kiyota T, Pokhrel RM, Chiaro G, Katagiri T, Sharma K and Wilkinson S (2015) The 2015 Gorkha Nepal earthquake: Insights from earthquake damage survey. *Frontiers in Built Environment* 1: 8.
- International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE) (1999) Manual for Zonation on Seismic Geotechnical Hazards (Revised version). Technical Committee for earthquake geotechnical engineering, TC4. *The Japanese Geotechnical Society, Tokyo*, March.
- Jalayer F, De Risi R and Manfredi G (2015) Bayesian Cloud Analysis: Efficient structural fragility assessment using linear regression. *Bulletin of Earthquake Engineering* 13(4): 1183–1203.
- Kotz S and Nadarajah S (2000). *Extreme value distributions: theory and applications*. London, UK: Imperial College Press World Scientific.

- Krivoruchko K and Gribov A (2014) Pragmatic Bayesian kriging for non-stationary and moderately non-Gaussian data. In: Pardo-Igúzquiza E, Guardiola-Albert C, Heredia J and Moreno-Merino L (eds) *Mathematics of Planet Earth*. Berlin; Heidelberg: Springer, pp. 61–64.
- McGowan SM, Jaiswal KS and Wald DJ (2017) Using structural damage statistics to derive macroseismic intensity within the Kathmandu valley for the 2015 M7. 8 Gorkha, Nepal earthquake. *Tectonophysics* 714–715: 158–172.
- McGuire RK (2004) *Seismic Hazard and Risk Analysis, Monograph*. Oakland, CA: Earthquake Engineering Research Institute.
- Machuca-Mory DF and Deutsch CV (2013) Non-stationary geostatistical modeling based on distance weighted statistics and distributions. *Mathematical Geosciences* 45: 31–48.
- Marache A, Breyse D, Piette C and Thierry P (2009) Geotechnical modeling at the city scale using statistical and geostatistical tools: The Pessac case (France). *Engineering Geology* 107(3–4): 67–76.
- Miano A, Jalayer F, De Risi R, Protà A and Manfredi G (2016) Model updating and seismic loss assessment for a portfolio of bridges. *Bulletin of Earthquake Engineering* 14(3): 699–719.
- Omre H (1987) Bayesian kriging—Merging observations and qualified guesses in kriging. *Mathematical Geology* 19(1): 25–39.
- Omre H and Halvorsen KB (1989) The Bayesian bridge between simple and universal kriging. *Mathematical Geology* 21(7): 767–786.
- Pilz J and Spöck G (2008) Why do we need and how should we implement Bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment* 22(5): 621–632.
- Pokhrel RM, Kuwano J and Tachibana S (2013) A kriging method of interpolation used to map liquefaction potential over alluvial ground. *Engineering Geology* 152(1): 26–37.
- Rahman MZ, Siddiqua S and Kamal AM (2018) Geology and topography based VS30 map for Sylhet City of Bangladesh. *Bulletin of Engineering Geology and the Environment* 88(5): 3069–3083.
- Rajaure S, Asimaki D, Thompson EM, Hough S, Martin S, Ampuero JP, Dhital MR, Inbal A, Takai N, Shigefuji M, Bijukchhen S, Ichiyanaagi M, Sasatani T and Paudel L (2017) Characterising the Kathmandu valley sediment response through strong motion recordings of the 2015 Gorkha earthquake sequence. *Tectonophysics* 714–715: 146–157.
- Shrestha OM, Kolrala A, Karmacharya SL, Pradhananga UB, Pradhan PM and Karmacharya R (1998) *Engineering and Environmental Geological Map of the Kathmandu Valley, Scale 1:50,000*. Kathmandu: Department of Mines and Geology, Lainchaur.
- Stewart JP, Klimis N, Savvaidis A, Theodoulidis N, Zargli E, Athanasopoulos G, Pelekis P, Mylonakis G and Margaritis B (2014) Compilation of a local VS profile database and its application for inference of VS 30 from geologic-and terrain-based proxies. *Bulletin of the Seismological Society of America* 104(6): 2827–2841.
- Tallett-Williams S, Gosh B, Wilkinson S, Fenton C, Burton P, Whitworth M, Datla S, Franco G, Trieu A, Dejong M, Novellis V, White T and Lloyd T (2016) Site amplification in the Kathmandu Valley during the 2015 M7.6 Gorkha, Nepal earthquake. *Bulletin of Earthquake Engineering* 14(12): 3301–3315.
- Thompson EM, Baise LG, Kayen RE, Tanaka Y and Tanaka H (2010) A geostatistical approach to mapping site response spectral amplifications. *Engineering Geology* 114(3–4): 330–342.
- Thompson EM, Wald DJ and Worden CB (2014) A VS30 map for California with geologic and topographic constraints. *Bulletin of the Seismological Society of America* 104(5): 2313–2321.
- Wald DJ and Allen TI (2007) Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America* 97(5): 1379–1395.
- Webster R and Oliver MA (2007) *Geostatistics for Environmental Scientists*. 2nd ed. Chichester: John Wiley & Sons, pp. 315.
- Worden CB, Wald DJ, Allen TI, Lin K, Garcia D and Cua G (2010) A revised ground-motion and intensity interpolation scheme for ShakeMap. *Bulletin of the Seismological Society of America* 100(6): 3083–3096.
- Yoshida M and Igarashi Y (1984) Neogene to Quaternary lacustrine sediments in the Kathmandu Valley, Nepal. *Journal of Nepal Geological Society* 4(Special Issue): 73–100.